Received: April 1, 2021. Revised: April 20, 2021.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Detection of Frame Duplication Using Multi Scale Local Oriented Feature Descriptors

Girish Nagaraj[1]* **Nandini Channegowda[2]**

[1]*Department of Information Science and Engineering,*
Dayananda Sagar Academy of Technology and Management, Bengaluru, India
[2]*Department of Computer science and Engineering,*
Dayananda Sagar Academy of Technology and Management, Bengaluru, India
* Corresponding author's Email: Girish.pt.6@gmail.com

Abstract: In recent decades, frame duplication is a common inter frame tampering operation in the digital videos. To find the duplicate frames with better computational time, multi scale local oriented feature descriptors are proposed in this paper. Initially, histogram equalization is used to improve the visual quality of the images or videos, which are collected from surrey university library for forensic analysis dataset. Then, feature extraction is accomplished utilizing binary robust invariant scalable keypoint, speeded up robust features, and maximally stable extremal regions to extract feature vectors or key points from the enhanced images. After identifying the keypoints, matched keypoints are evaluated by hamming distance and k-means clustering algorithm from the source and moving video frames. The multi scale local oriented feature descriptors with two step feature matching significantly decreases the computational time and effectively determines forged and non-forged frames in the video sequences. Simulation results showed that the proposed model achieved better performance in passive video forgery detection in terms of accuracy, sensitivity, f-score and specificity. Compared to the existing approaches like spatio temporal context learning, inter-frame forgery detection algorithm and adaptive parameter-based visual background extractor algorithm, the proposed model obtained better detection accuracy of 95.10%, and average detection time of 1.04 seconds per frame.

Keywords: Binary robust invariant scalable keypoint, Hamming distance, Histogram equalization, K-means clustering, Passive video forgery detection, Speeded up robust features.

1. Introduction

In recent times, the rapid growth of multimedia technology and user friendly editing software's like mokey by imaginer systems, and Photoshop and premiere by adobe collects and easily tamper the videos [1]. Generally, video tampering greatly affects the original video sequences and mislead the audiences, so an effective technology is needed to determine the authenticity of a video [2, 3]. In recent period, the multimedia forensics field is emerged to authenticate the veracity and integrity of videos or images [4]. The video forensics activity is categorized into two types such as passive forensics and active forensics. In active forensics, validation information is used for authentication while generating the video sequences, where the active approaches are limited in applicability [5]. In passive forensics, the integrity, and veracity of a video sequence is authenticated without using validation information, where the passive approaches are effective in practical applications [6]. In the field of multimedia security community, passive forensics is a growing research area on digital images [7]. Over the past few decades, many passive approaches are developed by the researchers, which are majorly classified into four types such as geometric-based, camera-based, pixel-based, and format-based passive approaches [8, 9].

In passive video forgery detection, computational time is the major problem, since the videos consists of thousands of frames. Several conventional approaches are developed by the researchers to detect

frame duplication forgeries. In this research study, a new model is proposed to perform better passive video forgery detection. At first, the input video sequences are collected from Surrey University Library for Forensic Analysis (SULFA) dataset. Then, the visual quality of the collected images is improved using histogram equalization method. Further, the keypoints are determined utilizing multi scale local oriented feature descriptors such as Binary Robust Invariant Scalable Keypoint (BRISK), Speeded up Robust Features (SURF), and Maximally Stable Extremal Regions (MSER). Additionally, the key points are matched by using hamming distance measure and k-means clustering algorithm. In this scenario, feature matching is carried out with individual clusters, which significantly decreases the time taken for matching the features. Finally, matched keypoints are normalized to the range of 0 to 1, if the threshold value is >0.3, the respective frame is considered as a forged, or else it is a non-forged frame. In the experimental section, the proposed model performance is validated by means of accuracy, sensitivity, f-score, and specificity.

This research paper is prepared as follows; some recent papers on the topic "passive video forgery detection" are surveyed in the Section 2. The proposed model is briefly explained in the Section 3 with proper mathematical expressions. The experimental investigations about the proposed model is represented in the Section 4. Section 5 indicates the conclusion of present research work.

2. Related works

Singh, and Singh [10] developed two approaches to detect duplication forgeries in the video sequences. By obtaining the mean features, approach I detects three dissimilar forms of copy moved frame duplication in the videos to evaluate correlation between the sequences. Approach II detects the copy moved duplication forgeries by locating the error positions with threshold values in order to compute the similarity between two frames or affected frames. The experimental results showed that the developed approaches achieved higher execution time efficiency, and detection accuracy in passive video forgery detection compared to the latest approaches on SULFA dataset. However, the developed approaches cannot detect the tampered areas that were subjected to mirror operation.

Zhao [11] developed a new video based passive blind forensic method to recognize inter frame forgeries based on similarity analysis. In this literature, the developed method includes two steps such as feature extraction and matching. In the initial step, feature extraction was carried out using SURF descriptor, and Hue Saturation Value (HSV) color histogram comparison. Then, Fast Library for Approximate Nearest Neighbors (FLANN) was applied for feature matching. In this literature study, S-V color and H-S color histograms were determined for each frame in a video shot and then calculate the similarity between histogram values to locate and detect the tempered frames in a video shot. Additionally, SURF feature descriptor and FLANN matching were applied to improve the performance of forgery detection in the passive videos. The experimental results demonstrated that the developed method was precise and efficient in light of forgery localization, and identification. However, the developed method obtained lower efficiency for a tampered video with large static scene.

Su, and Li [12] developed a novel approach to detect and locate forgeries in the video sequences and then summarize the characteristics of duplication forgeries. Initially, the developed approach extracts the feature vectors or keypoints in the video frames, and then search the tempered areas in the present frame. By utilizing spatio temporal context learning, the developed approach detects the tempered areas in the residual frames. Experimental results showed that the developed approach achieved better performance in passive video forgery detection compared to the existing methods in light of execution time and detection accuracy. As a future enhancement, the developed approach needs to concentrate on identifying smaller region to further enhance the forgery detection performance.

Kharat, and Chougule [13] developed a two-step approach to identify suspicious frames in the video sequences. In this literature study, SIFT descriptor was used to extract feature vectors or interested keypoints from the frames and then compared the keypoints with other frames to take the decision. Lastly, random sample consensus approach was applied to detect and locate duplicate frames. In this literature study, the developed two-step approach was tested on uncompressed and compressed videos with variable compression rate. Experimental result showed that the developed two-step approach obtained better accuracy in detecting the tampered frames compared to the prior approaches. However, the developed two-step approach unable to locate the spatial regions, which were altered that was considered as a major concern in this study.

D'Avino [14] developed a new deep learning model for video forgery detection on the basis of recurrent neural networks and auto-encoder. Initially, auto-encoder learns an intrinsic model of the source on a few pristine video frames. If the frames do not

fit to the learned model, where it is encoded with a higher reconstruction error and the forged materials were singled out as anomalous. Further, the temporal dependencies were exploited by implementing long short term memory model. Extensive experiments showed that the developed deep learning model achieved better performance in passive video forgery detection, especially on the compressed videos, which were downloaded from YouTube. While performing experiments with deep learning model, large set of input samples were required to achieve comparable performance, and it was computationally expensive.

Fadl [15] developed an effective approach using statistical textural features to locate and detect inter frame duplication forgery in the passive videos. Initially, the video sequences were divided into shots based on edge change ratio. Then, true positive was computed for each shot, and Gray-Level Co-Occurrence Matrix (GLCM) features were extracted in order to represent feature vectors. Lastly, the correlation between the identical feature vectors were determined to eliminate false positives. The simulation results showed that the developed approach obtained high accuracy on frame duplication or forgery detection with lower computational time. However, the developed approach needs to enhance shot boundary detection to obtain high accuracy, because incorrect shot increases the false positives.

Li [16] developed an inter-frame forgery detection scheme based on 2D phase congruency and k-means clustering algorithm for surveillance video. Firstly, compute 2D phase congruency for each frame. Then, correlation coefficients of adjacent frames and the variation of consecutive correlation coefficients were obtained. Finally, the discontinuous points caused by tampering were detected using k-means clustering algorithm. Experimental results showed that developed approach effectively detect, and localize the tampering positions. As a future improvement, it is essential to find a better solution for improving the precision of detecting frame deletion.

Su [17] developed a forgery detection algorithm for detecting video foreground removal. The developed algorithm initially computes the energy factor of every video frame for identifying the forged frames. Next, an adaptive parameter based visual background extractor algorithm was developed for detecting the suspected regions from the forged video frames. Then, calculate the difference of the energy factor between forged frames and authentic frames in order to eliminate the false detection, and finally locate to the tampering traces. Simulation results showed that the developed algorithm obtained better performance in terms of classification accuracy and computational efficiency. However, the developed forgery detection algorithm performance is degraded under complex backgrounds like water ripples, brightness change, slightly shaking screens, noise and swaying trees. In order to address the aforementioned problems, a new model is proposed in this research paper to improve the performance of passive video forgery detection.

3. Methodology

In recent decades, passive video forgery detection is an emerging research topic in the field of multimedia security community. In this research study, a new model is proposed to locate the duplicate regions in the videos without affecting the video quality and provides better performance by means of accuracy, sensitivity, specificity, and f-score. In passive video forgery detection, the proposed model includes following steps like **data collection:** SULFA dataset, **data pre-processing:** histogram equalization, feature extraction: BRISK, SURF, and MSER, and passive video forgery detection: feature matching by hamming distance with k-means clustering. Workflow of proposed model is given in Fig. 1.

3.1 Data collection

In this study, the input video sequences are collected from SULFA dataset to evaluate the

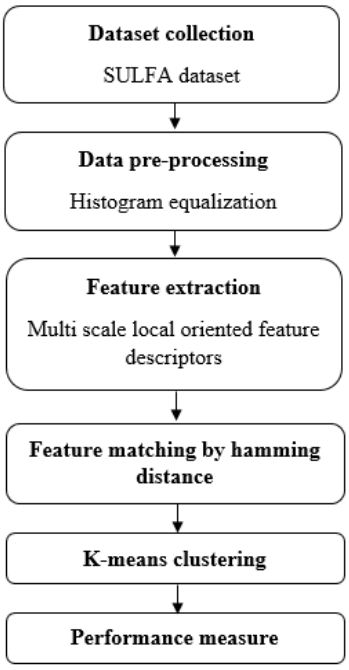

Figure. 1 Work flow of proposed model

proposed model performance. SULFA dataset comprises of original and forged video files, which are publicly available in University of Surrey's website [18]. The SULFA dataset comprises of 20 video sequences (10 originals and 10 forged), which are collected from three camera sources; Fujifilm S2800HD, Canon SX220, and Nikon S3000. The time duration of every video sequences is ten seconds with the pixel resolution of 320×240, and includes thirty video frames per second. In SULFA dataset, video sequences are shot by considering spatial and temporal properties [19]. Additionally, the video sequences contain simple and complex scenes with and without utilizing camera support in order to present life time scenarios. The sample video frames of SULFA dataset is represented in Fig. 2.

Dataset link:
https://sites.google.com/site/rewindpolimi/downloads/datasets/video-copy-move-forgeries-dataset

3.2 Data pre-processing

After data collection, the video sequences are converted into frames, and then the RGB images are converted into grayscale images to ease the representation of collected data. Further, histogram equalization method is applied to enhance the visibility level of the images by increasing the global image contrast [20]. Let us consider a gray scale image x and n_i is indicated as number of gray level occurrences i. The occurrence probability of image pixel value is calculated by using Eq. (1).

$$p_x(i) = \frac{n_i}{n}, 0 \le i < L \tag{1}$$

where, n is indicated as total image pixels, L is represented as number of image gray levels (256), and $p_x(i)$ is denoted as histogram value of image pixels i, which is normalized to [0, 1]. Then, the Cumulative Distribution function (CDF) is determined for p_x using Eq. (2).

$$cdf_x(i) = \sum_{j=0}^{i} p_x(x = j) \tag{2}$$

Further, develop a transformation form $y = T(x)$ to create a new image y with flat histogram value. The transformed images have linearized CDF that is mathematically defined in the Eq. (3) and (4).

$$cdf_y(i) = iK \tag{3}$$

$$cdf_y(y') = cdf_y(T(k)) = cdf_x(k) \tag{4}$$

where, K *and* T are denoted as constant values, which ranges between [0, 1], and k is in the range of [0, L]. Lastly, a simple transformation is applied to map the pixel values back into their original image, which is mathematically denoted in Eq. (5). After data pre-processing, multi scale local oriented feature descriptors are applied on the pre-processed image y' to extract the feature vectors. The sample pre-processed image is given in Fig. 3.

$$y' = y \times (max(x) - min(x)) + min(x) \tag{5}$$

Figure. 2 Sample video frames of SULFA dataset

(a) (b) (c)

Figure. 3: (a) source frame, (b) converting source frame into grayscale image, and (c) Enhanced frame

3.3 Feature extraction

After enhancing the quality of images, feature extraction is accomplished using multi scale local oriented feature descriptors like BRISK [21], SURF [22], and MSER [23] in order to extract feature vectors from the images. Feature extraction is the procedure by which the collected data is reduced by determining the key features in data for machine learning.

BRISK: It is a method for binary description and scale space key point detection. In BRISK feature descriptor, keypoints are detected in the octave layers of image pyramid. In this method, scale and the location of each keypoint is transformed into continuous domain representation using quadratic function fitting. BRISK descriptor is calculated as a binary string once the BRISK feature vectors are determined. BRISK descriptor includes two phases; (i) compute keypoints orientation to create a rotation invariant descriptor, (ii) evaluate image brightness to capture the local region properties.

SURF: This descriptor uses a BLOB detector to determine the interested keypoints on the basis of Hessian matrix. SURF feature descriptor utilizes wavelet responses in both vertical and horizontal directions using adequate Gaussian weights for orientation assignment. The neighbourhood regions around the interested keypoints are selected and then the wavelet response is considered for every region. In addition, sign of Laplacian is computed in the selected regions for distinguishing bright blobs from the dark backgrounds.

MSER: It is invariant to image intensity and affine transformation that ensures the regions, which are extracted while the illumination is changed and the tempered areas are transformed. In MSERs, the computational complexity is linear, and also it is stable over a wide range of thresholds.

3.4 Passive video forgery detection

After extracting the feature vectors or keypoints, feature matching is accomplished using hamming distance measure [24] in order to decrease the false positives. The step by step procedure of feature matching is given below,

Step 1: After extracting the keypoints in denoised images y' (source and moving frames), feature matching is carried out using hamming distance. The keypoints are extracted using BRISK, SURF, and MSER feature descriptors, and the extracted keypoints are described by binary hamming code.

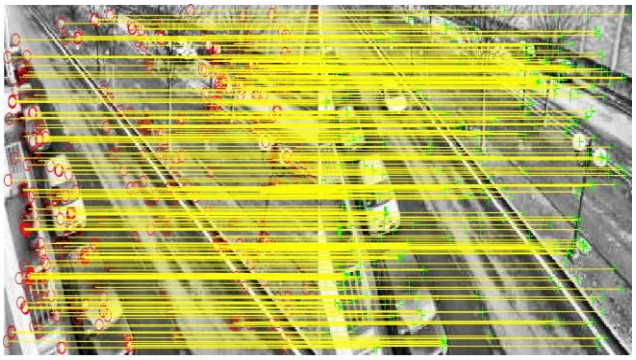

Figure. 4 Feature matching between source frame and moving frame

Step 2: Then, nearest adjacent neighbour approach of hamming distance measure is utilized to make initial keypoints matching. Further, the matched keypoints G are arranged in the ascending order based on the hamming distance t, $G = \{(G_1, G_1'), (G_2, G_2'), \ldots., (G_n, G_n')\}$, where n is represented as number of matched keypoints.

Step 3: Initially, select three pairs of matching points $(G_1, G_1'), (G_2, G_2')\ and\ (G_3, G_3')$ and evaluate whether the selected keypoints meet the similar sequence structure or not. If the condition is satisfied, choose 1st two pairs of keypoints $(G_1, G_1'), and\ (G_2, G_2')$ as the reference points. If the condition is not satisfied, choose three adjacent keypoints from the 2nd pair of matching keypoints, and evaluate whether the keypoints meet the similar sequence structure or not. Further, make a test loop until the three pairs of adjacent keypoints meet the condition, and choose top two pairs as the reference keypoints $(G_{q-1}, G_{q-1}'), and\ (G_q, G_q')$.

Step 4: Choose a pair of matching keypoints (G_m, G_m') from the residual matching points. Evaluate whether the selected keypoint $(G_{q-1}, G_{q-1}'), (G_q, G_q')\ and (G_m, G_m')$ meet the sequence structure or not. If the condition is not satisfied, eliminate (G_m, G_m'), orelse retain the matching keypoint. After verifying all the keypoints in sample G, and then generate a new sample U. The feature matching is graphically represented in Fig. 4.

Step 5: After generating the new sample U, k-means clustering [25] is used to classify the retained keypoints into two categories like tampered region W' and source region Z'. Any pair of matching keypoints w_n and z_n will be eliminated, if the keypoints meet either of the two conditions, as stated in Eq. (6).

$$z_n \in Z'\ and\ w_n \in Z', \text{ or}$$
$$z_n \in W'\ \ and\ w_n \in W' \quad (6)$$

(a) (b)

Figure. 5: (a) non-forged frame and (b) forged frame

The number of retained matching point pairs are computed after removing the mismatching keypoints. The number of matched keypoints are normalized into the range 0 to 1. If the normalized threshold value is >0.3, the frame is considered as forged and if the normalized threshold value is <0.3, the frame is considered as non-forged. Sample frame of forged and non-forged is represented in Fig. 5.

4. Experimental result

In the experimental section, the proposed model performance is tested on a publicly available dataset named SULFA. The test videos used in this research paper are MPEG 2 encoded with frame rate of 30 per seconds. In this research study, the proposed model performance is simulated by using MATLAB (2018a) software with the system requirements; operating system: Microsoft windows 10, memory size: 16 GB, processor: Intel(R) Core (TM) 2 i7-4700MQ 2.4GHz, and video card: NVIDIA GeForce GT 755M. In this study, the effectiveness of the proposed model performance is evaluated by comparing with the benchmark approaches such as spatio temporal context learning [12], inter-frame forgery detection algorithm [16], and adaptive parameter-based visual background extractor algorithm [17] on SULFA dataset. In addition, the proposed model performance is validated in light of detection accuracy, sensitivity, specificity and f-score.

Accuracy is an important performance measure in image processing application, where it is a ratio of correctly predicted class from total testing class. Accuracy assessment is mathematically defined in Eq. (7).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \quad (7)$$

Additionally, sensitivity estimates the proportion of positives, which are correctly identified. Specificity estimates the proportion of negatives, which are correctly identified. The f-score is a harmonic mean of precision and recall, where sensitivity, specificity and f-score are mathematically represented in the Eq. (8), (9), and (10).

$$Sensitivity = \frac{TP}{FN + TP} \times 100 \quad (8)$$

$$Specificity = \frac{TN}{FP + TN} \times 100 \quad (9)$$

$$F - score = \frac{2TP}{2TP + FP + FN} \times 100 \quad (10)$$

where, true positive is represented as TP, true negative is denoted as TN, false positive is indicated as FP, and false negative is stated as FN.

4.1 Quantitative analysis

In this section, the proposed model performance is investigated on SULFA dataset by means of sensitivity, specificity, accuracy and f-score. The SULFA dataset consists of 20 videos (10 original videos and 10 forged videos) with the resolution of 320×240. In this scenario, the performance is validated for all 10 video sequences with individual and hybrid feature extraction. By investigating the Table 1 and 2, the hybrid feature extraction achieved better sensitivity, and specificity value in all 10 video sequences. The proposed model obtained mean sensitivity value of 70.04% and mean specificity value of 67.86% in passive video forgery detection. Further the graphical comparison of the proposed model in terms of sensitivity, and specificity is represented in Fig. 6 and 7.

In recent periods, the passive video forgery detection is a challenging task, because the tampered and source regions are highly similar. To highlight

Table 1. Performance analysis of the proposed model in terms of sensitivity

Sensitivity (%)				
Video sequences	**SURF**	**BRISK**	**MSER**	**Hybrid**
Video 1	67.86	78.69	68.29	**88.1**
Video 2	53.73	49.68	48.71	**58.42**
Video 3	53.01	49.23	48.05	**77.66**
Video 4	60.71	62.38	58.7	**70.28**
Video 5	63.67	59.59	55.25	**87.16**
Video 6	64.26	58.76	56.34	**80.59**
Video 7	63.05	71.41	61.01	**85.82**
Video 8	69.35	77.27	67.88	**84.69**
Video 9	62.52	69.5	62.14	**83.73**
Video 10	55.44	51.64	50.76	**64**
Mean	61.36	62.815	57.713	**78.045**

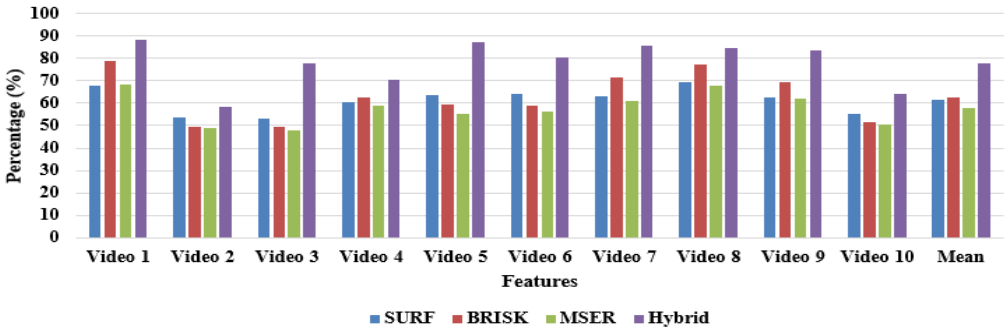

Figure. 6 Graphical comparison of the proposed model in terms of sensitivity

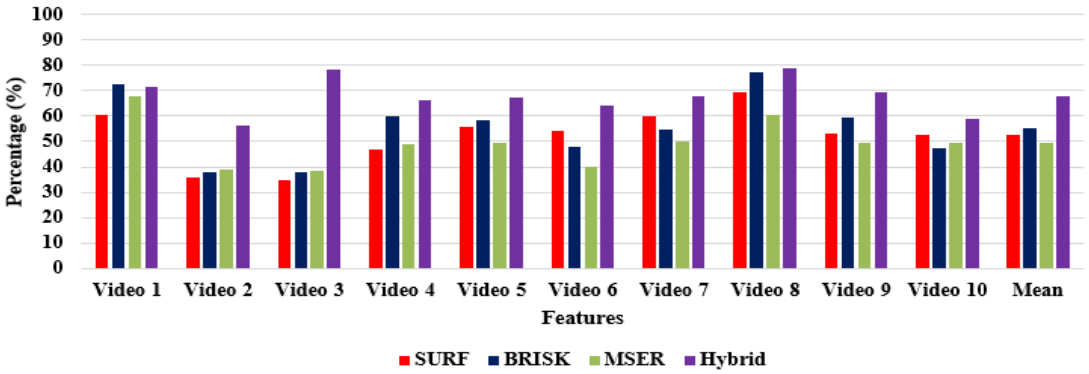

Figure. 7 Graphical comparison of the proposed model in terms of specificity

Table 2. Performance analysis of the proposed model in terms of specificity

Specificity (%)				
Video sequences	**SURF**	**BRISK**	**MSER**	**Hybrid**
Video 1	60.53	72.65	67.93	**71.54**
Video 2	35.78	38.11	39.23	**56.04**
Video 3	34.97	38.08	38.45	**78.47**
Video 4	46.96	59.71	49.11	**66.19**
Video 5	55.67	58.6	49.59	**67.44**
Video 6	54.39	48.08	40.2	**64.3**
Video 7	59.99	54.82	49.75	**67.81**
Video 8	69.33	77.13	60.58	**78.69**
Video 9	53.36	59.22	49.53	**69.23**
Video 10	52.78	47.47	49.39	**58.95**
Mean	52.376	55.387	49.376	**67.866**

Table 3. Performance analysis of the proposed model in terms of detection accuracy

Accuracy (%)				
Video sequences	**SURF**	**BRISK**	**MSER**	**Hybrid**
Video 1	87.02	87.56	75.12	**97.51**
Video 2	94.77	88.31	87.08	**95.64**
Video 3	93.96	87.57	86.5	**95.82**
Video 4	95.25	93.04	90.82	**96.46**
Video 5	91.57	87.26	84.36	**84.36**
Video 6	92.34	84.29	84.29	**93.87**
Video 7	91.24	85.64	82.97	**94.16**
Video 8	94.46	87.18	87.55	**94.6**
Video 9	91.29	86.62	83.54	**95.1**
Video 10	91.06	81.7	81.28	**93.07**
Mean	92.296	86.917	84.351	**94.059**

this concern, several algorithms are developed by the researchers, which are majorly high dimensional and computationally complex, particularly while analysing the videos with high resolution. In this research study, multi scale local oriented feature descriptors; BRISK, SURF, and MSER are employed to retain mirror invariance that significantly diminishes the feature dimensionality and retain the characteristics of scale invariance and rotation of a frame to improve the performance of video forgery detection.

Similarly, in the Tables 3 and 4, the proposed model performance is validated on SULFA dataset in terms of detection accuracy and f-score. In this scenario, the hybrid feature extraction in combination with feature matching by hamming distance and k-means clustering obtained a mean detection accuracy of 95.10% and mean f-score value of 69.81%, which are effective compared to individual feature extraction. In the Table 3 and 4, the proposed model performance is validated for all 10 video sequences in terms of detection accuracy and f-score.

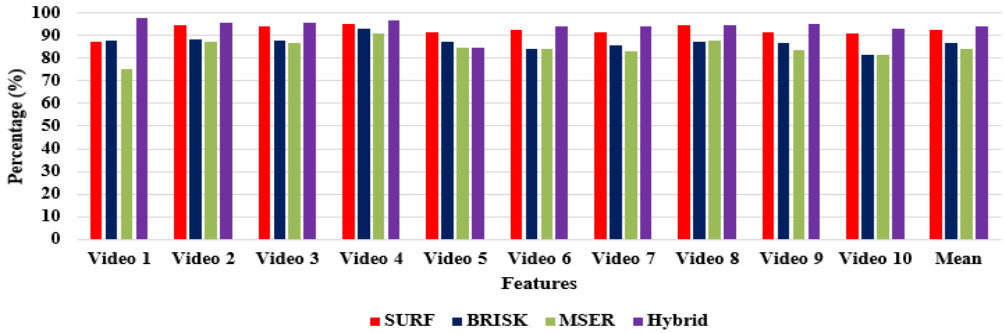

Figure. 8 Graphical comparison of the proposed model in terms of accuracy

Table 4. Performance analysis of the proposed model in terms of f-score

F-Score (%)				
Video sequences	**SURF**	**BRISK**	**MSER**	**Hybrid**
Video 1	72.56	82.42	68.29	92.56
Video 2	53.92	49.39	48.81	55.13
Video 3	53.84	49.13	48.67	78.57
Video 4	63.06	63.79	60.38	66.74
Video 5	64.19	60.23	56.1	73.64
Video 6	62.26	55.45	55.45	65.06
Video 7	63.05	66.12	60.83	71.18
Video 8	62.85	59.26	59.64	62.1
Video 9	64.34	68.17	63.46	75.02
Video 10	55.64	51.02	50.76	58.19
Mean	61.571	60.498	57.239	69.819

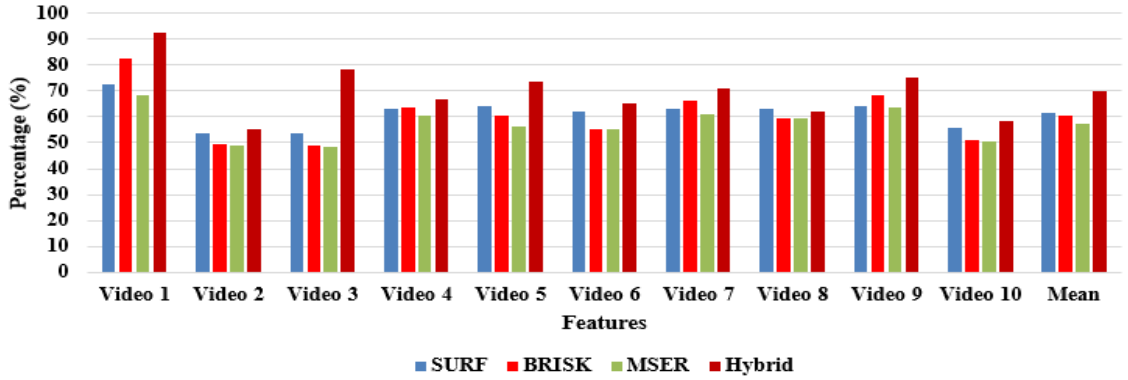

Figure. 9 Graphical comparison of the proposed model in terms of f-score

In this study, two steps; feature matching by hamming distance and k-means clustering are developed to decrease the number of mismatched keypoints that improves the detection accuracy in passive video forgery detection. In k-means clustering, matching step is done with each cluster individually that decreases the time needed for feature matching. In addition, the computational time of BRISK is 1.08 seconds per frame, SURF is 1.07 seconds per frame, MSER is 1.08 seconds per frame, and the hybrid descriptors consumes 1.04 seconds per frame. The graphical comparison of the proposed model in terms of accuracy and f-score is denoted in the Figs. 8 and 9.

4.2 Comparative analysis

In this section, comparative analysis between the proposed model and the existing approach is denoted in Table 5. L. Su, and C. Li [12] implemented a novel approach to detect and locate forgeries in the videos and to summarize the properties of duplication forgeries. At first, key points in the video frames were extracted using Mirror and Inversion invariant generalization for SIFT (MI-SIFT) descriptor. Then, the tampered areas in the video frames were located by utilizing spatio temporal context learning. In the experimental section, the developed novel passive

Table 5. Comparative analysis between the proposed and the existing approach

Methodology	**Detection accuracy (%)**	**Detection time (seconds)**
Spatio temporal context learning [12]	92.6	-
Inter-frame forgery detection algorithm [16]	93.75	-
Adaptive parameter-based visual background extractor algorithm [17]	86.58	2.69
Multi scale local oriented feature descriptors	**95.10**	**1.04**

forgery detection approach performance was validated on SULFA dataset by means of detection accuracy. The extensive experiment showed that the developed passive forgery detection approach achieved 92.6% of detection accuracy. Additionally, Li, [16] developed an inter-frame forgery detection algorithm, which composed of feature extraction, and abnormal point localization. In feature extraction step, extract the 2-D phase congruency of each frame. Further, compute the correlation between the adjacent frames, and then the abnormal points were detected using k-means clustering algorithm. Lastly, normal, and abnormal points were clustered into two categories. The experimental result demonstrated that the inter-frame forgery detection algorithm achieved 93.75% of detection accuracy.

Su [17] developed forgery detection algorithm that detects video foreground removal. Initially, the developed algorithm computes energy factor for each frame for identifying the forged frames in the video sequences. Then, an adaptive parameter based visual background extractor algorithm was applied to detect the suspected regions in the forged frames. Lastly, the difference of the energy factor between the suspected regions in the forged and authentic video frames was calculated for eliminating the false detection. Hence, the experimental analysis showed that the developed algorithm achieved 86.58% of accuracy, and average detection time of 2.69 seconds per frame.

Compared to the existing algorithms, the proposed model obtained 95.1% of accuracy, which is better in passive video forgery detection. Further the proposed model consumes limited computational time of 1.04 seconds per frame and it achieved better performance in passive video forgery detection compared to the literatures [12, 16, 17].

In this research paper, a new model is proposed to perform passive video forgery detection. In this study, hybrid feature extraction in combination with feature matching by hamming distance measure and k-means clustering algorithm retains the properties of mirror invariance, rotation and scale invariance to achieve effective performance in frame duplication or forgery detection. In this study, the proposed model achieved better performance in passive video forgery detection in light of f-score, accuracy, sensitivity and specificity. In the comparative section, proposed model showed maximum of 2.50% improvement in detection accuracy compared to the existing novel passive forgery detection approaches. Additionally, the overall computational time of proposed model is 345 seconds or 1.04 seconds per frame, which is better compared to individual feature extraction. In the future work, an improved clustering algorithm can be included in the proposed model to further improve the performance of passive video forgery detection.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision, and project administration, have been done by 2nd author.

References

[1] K. N. Sowmya, H. R. Chennamma, and L. Rangarajan, "Video authentication using spatio temporal relationship for tampering detection", *Journal of Information Security and Applications*, Vol. 41, pp. 159-169, 2018.

[2] H. Kaur and N. Jindal, "Deep Convolutional Neural Network for Graphics Forgery Detection in Video", *Wireless Personal Communications*, pp. 1-19, 2020.

[3] S. Jia, Z. Xu, H. Wang, C. Feng, and T. Wang, "Coarse-to-fine copy-move forgery detection for video forensics", *IEEE Access*, Vol. 6, pp. 25323-25335, 2018.

[4] P. He, X. Jiang, T. Sun, S. Wang, B. Li, and Y. Dong, "Frame-wise detection of relocated I-frames in double compressed H. 264 videos based on convolutional neural network", *Journal of Visual Communication and Image Representation*, Vol. 48, pp. 149-158, 2017.

[5] L. Su, C. Li, Y. Lai, and J. Yang, "A fast forgery detection algorithm based on exponential-Fourier moments for video region duplication", *IEEE Transactions on Multimedia*, Vol. 20, No. 4, pp. 825-840, 2017.

[6] S. Fadl, Q. Han, and Q. Li, "CNN spatiotemporal features and fusion for surveillance video forgery detection", *Signal Processing: Image Communication*, Vol. 90, pp. 116066, 2021.

[7] M. Aloraini, M. Sharifzadeh, and D. Schonfeld, "Sequential and Patch Analyses for Object Removal Video Forgery Detection and Localization", *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[8] M. Saddique, K. Asghar, U. I. Bajwa, M. Hussain, and Z. Habib, "Spatial video forgery detection and localization using texture analysis of consecutive frames", *Advances in Electrical and Computer Engineering*, Vol. 19, No. 3, pp. 97-108, 2019.

[9] G. Chittapur, S. Murali, and B. S. Anami, "Copy Create Video Forgery Detection Techniques Using Frame Correlation Difference by Referring SVM Classifier", *International Journal of Computer Engineering In Research Trends*, Vol. 6, No. 12, pp. 2349-7084, 2019.

[10] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation", *Multimedia Tools and Applications*, Vol. 78, No. 9, pp. 11527-11562, 2019.

[11] D. N. Zhao, R. K. Wang, and Z. M. Lu, "Inter-frame passive-blind forgery detection for video shot based on similarity analysis", *Multimedia Tools and Applications*, Vol. 77, No. 19, pp. 25389-25408, 2018.

[12] L. Su and C. Li, "A novel passive forgery detection algorithm for video region duplication", *Multidimensional Systems and Signal Processing*, Vol. 29, No. 3, pp. 1173-1190, 2018.

[13] J. Kharat and S. Chougule, "A passive blind forgery detection technique to identify frame duplication attack", *Multimedia Tools and Applications*, pp. 1-17, 2020.

[14] D. D'Avino, D. Cozzolino, G. Poggi, and L. Verdoliva, "Autoencoder with recurrent neural networks for video forgery detection", *Electronic Imaging*, Vol. 2017, No. 7, pp. 92-99, 2017.

[15] S. Fadl, A. Megahed, Q. Han, and L. Qiong, "Frame duplication and shuffling forgery detection technique in surveillance videos based on temporal average and gray level co-occurrence matrix", *Multimedia Tools and Applications*, pp. 1-25, 2020.

[16] Q. Li, R. Wang, and D. Xu, "An inter-frame forgery detection Algorithm for surveillance video", *Information*, Vol. 9, No. 12, pp. 301, 2018.

[17] L. Su, H. Luo, and S. Wang, "A novel forgery detection algorithm for video foreground removal", *IEEE Access*, Vol. 7, pp. 109719-109728, 2019.

[18] G. Qadir, S. Yahaya, and A. T. Ho, "Surrey university library for forensic analysis (SULFA) of video content", *IET Conference on Image Processing*, 2012.

[19] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Local tampering detection in video sequences", In: *Proc. of IEEE 15th International workshop on multimedia signal processing, IEEE*, pp. 488-493, 2013.

[20] Y. Zhu and C. Huang, "An adaptive histogram equalization algorithm on the image gray level mapping", *Physics Procedia*, Vol. 25, pp. 601-608, 2012.

[21] P. Niyishaka and C. Bhagvati, "Copy-move forgery detection using image blobs and BRISK feature", *Multimedia Tools and Applications*, Vol. 79, No. 35, pp. 26045-26059, 2020.

[22] X. Bo, W. Junwen, L. Guangjie, and D. Yuewei, "Image copy-move forgery detection based on SURF", In: *Proc. of International Conference on Multimedia Information Networking and Security, IEEE*, pp. 889-892, 2010.

[23] Y. Zhu, X. Shen, and Y. Liu, "Copy-move forgery detection based on local intensity order pattern and maximally stable extremal regions", *Journal of Intelligent & Fuzzy Systems*, Vol. 37, No. 6, pp. 7761-7768, 2019.

[24] W. Chen, Y. Zhang, J. Wen, K. Li, and G. Yang, "An Application of Improved RANSAC Algorithm in Visual Positioning", In: *Proc. of 8th Joint International Information Technology and Artificial Intelligence Conference, IEEE*, pp. 1358-1362, 2019.

[25] O. M. Al-Qershi and B. E. Khoo, "Copy-move forgery detection using on locality sensitive hashing and k-means clustering", In: *Proc. Of Information Science and Applications, Springer, Singapore*, pp. 663-672, 2016.